基础研究

基于介电特性的人体恶性胃组织支持向量机辅助诊断方法

张 洒[1],厉 周[2],辛学刚[1]
[1]南方医科大学生物医学工程学院,广东 广州 510515;[2]南方医科大学珠江医院普外科,广东 广州 510280

摘要:目的 基于正常和恶性胃组织介电特性值的差异,运用支持向量机(SVM)对介电特性进行自动鉴别。**方法** 用开端同轴探头法测量正常和恶性胃组织在42.58~500 MHz频率范围内的介电特性,并对测得的介电特性数据进行Cole-Cole模型拟合。接收机操作特性(ROC)曲线分析法被用来对各频率点下介电常数、电导率和Cole-Cole拟合参数的鉴别能力进行评估。SVM被用来对正常和恶性胃组织进行鉴别,鉴别正确率由k折交叉验证进行计算。**结果** 在测量频率范围内,5个低端频率点下介电常数的ROC曲线下面积达到0.8以上。这5个频率下介电常数的组合作为特征值与SVM结合取得了最高鉴别正确率84.38%,MATLAB运行时间为3.40 s。**结论** 本文提出的基于介电特性的恶性人体胃组织支持向量机辅助诊断方法具有可行性。
关键词:介电特性;恶性胃组织;鉴别;ROC曲线;支持向量机

Support vector machine-assisted diagnosis of human malignant gastric tissues based on dielectric properties

ZHANG Sa[1], LI Zhou[2], XIN Xuegang[1]
[1]School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China; [2]Department of General Surgery, Zhujiang Hospital, SouthernMedical University, Guangzhou 510280, China

Abstract: Objective To achieve differential diagnosis of normal and malignant gastric tissues based on discrepancies in their dielectric properties using support vector machine. **Methods** The dielectric properties of normal and malignant gastric tissues at the frequency ranging from 42.58 to 500 MHz were measured by coaxial probe method, and the Cole-Cole model was used to fit the measured data. Receiver-operating characteristic (ROC) curve analysis was used to evaluate the discrimination capability with respect to permittivity, conductivity, and Cole-Cole fitting parameters. Support vector machine was used for discriminating normal and malignant gastric tissues, and the discrimination accuracy was calculated using k-fold cross-validation. **Results** The area under the ROC curve was above 0.8 for permittivity at the 5 frequencies at the lower end of the measured frequency range. The combination of the support vector machine with the permittivity at all these 5 frequencies combined achieved the highest discrimination accuracy of 84.38% with a MATLAB runtime of 3.40 s. **Conclusion** The support vector machine-assisted diagnosis is feasible for human malignant gastric tissues based on the dielectric properties.
Keywords: dielectric properties; malignant gastric tissues; discrimination; support vector machine

介电特性是人体组织的固有属性,包括介电常数 ε 和电导率 σ[1]。人体组织的介电特性值与组织自身的水含量、蛋白质含量及离子浓度等因素有关[2-4],与组织所处的微循环等微环境有关,并且具有频率依赖性。当组织发生癌变的时候,上述介电特性相关因素会发生改变,进而导致正常和恶性组织的介电特性值之间存在较为显著的差异[2, 4]。理论上,正常和恶性组织的介电特性值之间的差异,在组织形态学改变之前就已经存在,因此可能会成为早期发现癌变可疑部位的一种新的技术方法,发现可疑部位后再有针对性地实施活检,有可能实现癌症的早期诊断。

Gabriel等[5-7]系统地研究了健康人体组织的介电特性频率谱,并且建立了人体正常组织介电特性的数据库。同时对正常和恶性生物组织的介电特性进行测量,显示正常和恶性组织之间的介电特性的差异通常显著[8-12],这表明确实可能通过这种差异来实现组织良恶性鉴别。但是,目前还没有明确的标准指出,组织的介电特性差异达到多大,即可判断为可疑的癌变组织。因此,研究通过组织介电特性差异的程度来判断组织良恶性的判别方法具有重要的意义。为此,已经有学者开展了相关研究工作。其中,Truong等[13-14]在THz频率下,基于正常和恶性组织介电特性的差异,分别用机器学习的方法对乳腺和皮肤的正常和恶性组织进行鉴别,且对两

种组织的鉴别正确率最高均达到90%以上；Yilmaz等[15]在500 MHz~6 GHz频率范围内测量老鼠正常和恶性肝组织的介电特性，然后用机器学习的方法进行鉴别，并取得了高于90%的鉴别正确率。然而，目前还未见文献报导基于介电特性差异自动鉴别人体正常和恶性的胃组织的方法。本文采用开端同轴探头法在临床手术室中测量人体胃部正常和癌变组织在42.58~500 MHz频率范围内的介电特性，探索采用机器学习方法自动鉴别正常和恶性胃组织的可行性。利用统计分析方法中的接收机操作特性(ROC)曲线来判断采用多种参数对介电特性进行鉴别的能力，并从中选出鉴别能力较高的参数作为机器学习分类的特征值来完成基于机器学习方法的自动鉴别。鉴于本文的数据为数值型，且属于二分类问题，我们选择支持向量机(SVM)来对正常和恶性胃组织进行鉴别。

1 材料和方法

1.1 数据介绍

本研究采用开端同轴探头法[16-18]，在手术室测量新鲜切除的人体恶性胃组织及其周围正常组织的介电特性。测量的频率范围为42.58~500 MHz，共计43个频率点。共测量了来自26例病人的290个样本，其中包括100个正常样本和190个恶性样本，病人年龄在36~80岁之间，离体测量的温度范围为23.8~32.8 ℃。

本文也对测得的胃组织的介电特性进行Cole-Cole模型拟合[19-21]，下式是单极Cole-Cole拟合公式：

$$\varepsilon^*(\omega)=\varepsilon'(\omega)-j\varepsilon''(\omega)=\varepsilon_\infty+\frac{\Delta\varepsilon}{1+(j\omega\tau)^{1-a}}+\frac{\sigma_i}{j\omega\varepsilon_0}\quad(1)$$

其中，ω 是角频率，$\varepsilon^*(\omega)$ 是样本依赖频率的复介电常数数据，$\varepsilon'(\omega)$ 是依赖频率的介电常数，$\varepsilon''(\omega)$ 是依赖频率的介电损耗（可以转换为电导率：$\sigma(\omega)=\omega\varepsilon_0\varepsilon''(\omega)$），$\varepsilon_\infty$、$\sigma_i$、$\Delta\varepsilon$、$\tau$ 和 a 是根据实验数据估算出的Cole-Cole模型拟合参数。Cole-Cole拟合的具体操作是使用文献中提出的软件包来实现的[22]。图1显示的是两个不同样本(分别是正常和恶性组织)的原始测量数据及Cole-Cole拟合曲线。

这样，每个样本都包括43个频率下的介电常数和电导率，以及5个Cole-Cole拟合参数。这些介电特性

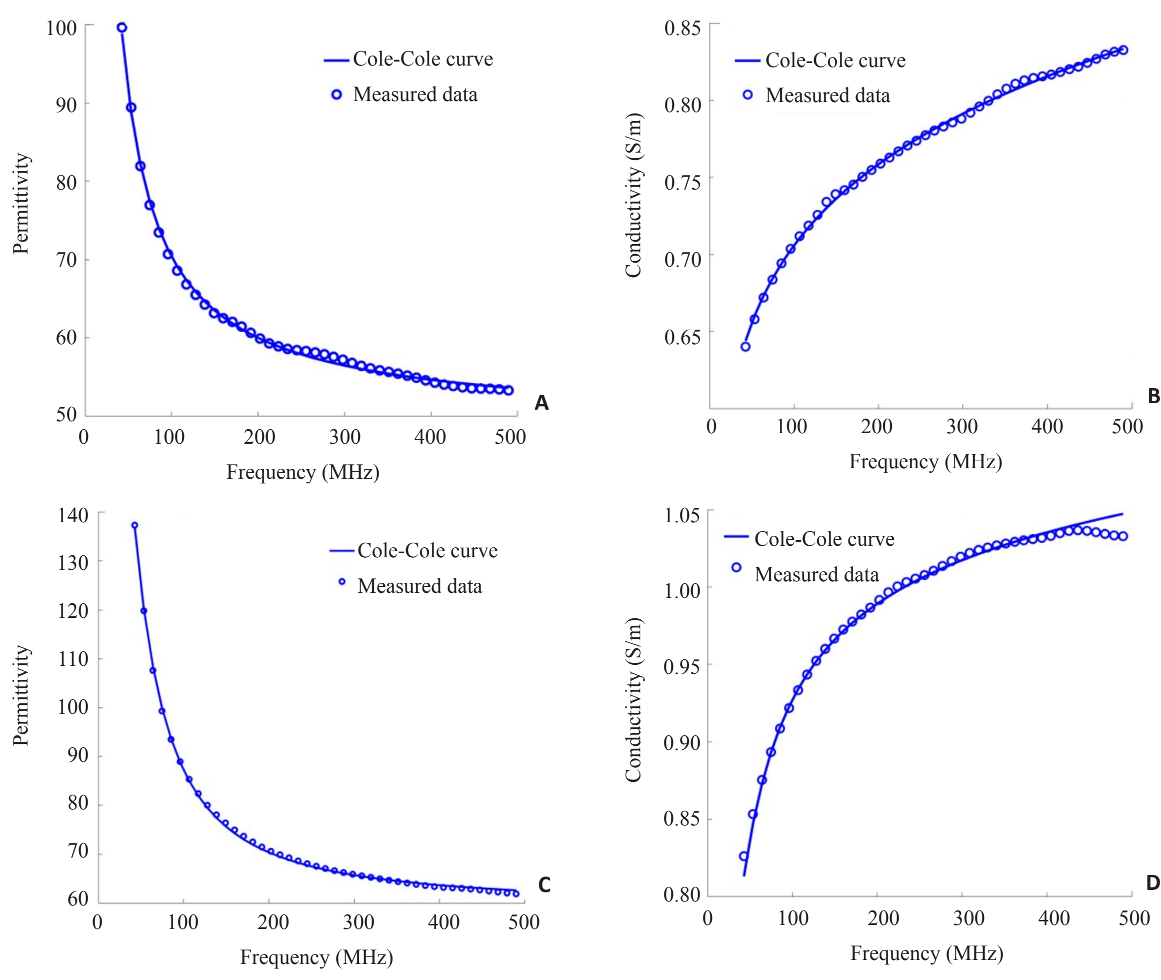


图1 Cole-Cole拟合结果图例

Fig.1 Example of Cole-Cole fitting results. *A*, *B*: Fitting graphs of permittivity and conductivity of normal tissue with frequency, respectively; *C*, *D*: Fitting graphs of permittivity and conductivity of malignant tissue with frequency, respectively.

值和拟合参数将用于下一步的ROC曲线分析，以判断其鉴别能力。

1.2 ROC曲线分析

ROC曲线[23-24]分析是组织分类器和可视化其性能的一种有用技术，通常用于医疗决策，近年来被来越来越多的在机器学习和数据挖掘领域使用。ROC曲线图是二维图，其横坐标代表假阳性率，纵坐标代表真阳性率，曲线上的每个点对应着一个判决阈值。作为ROC曲线分析的一个关键指标，ROC曲线下面积(AUC)不依赖任何阈值，其值在0.5~1之间。AUC越接近1，代表鉴别能力越高。

本研究运用ROC曲线来判断每一个测量频率下的介电常数、电导率，以及5个Cole-Cole拟合参数各自的鉴别能力，从而找到鉴别能力较高的特征值用于接下来的机器学习鉴别研究。

1.3 SVM介绍

SVM[25-28]是建立在统计学习理论的VC维理论和结构风险最小化原理基础上的一种机器学习方法。目前已有不少文献提到用SVM来鉴别正常和恶性组织[13-15]。本文运用二分类SVM，通过训练样本找到最优超平面，将正常和恶性胃组织分开在超平面的两侧，并利用该超平面对测试样本进行分类。

$X=(x_1,x_2,\ldots,x_L)^T$ 代表训练样本集，其中L是样本个数，$x_l \in R^M (l=1,2,\ldots,L)$ 表示一个样本，即特征向量。$i \in \{-1,1\}$，其中 $i=1$ 代表恶性组织，$i=-1$ 代表正常组织。由支持向量决定的超平面 $w \cdot x+b=1$ 和 $w \cdot x+b=-1$ 之间的距离为 $2/\|w\|$，求 $2/\|w\|$ 的最大值即求 $\|w\|^2$ 的最小值。这样，找最优超平面可通过最小化 $\|w\|^2$ 来进行。另外，训练样本中可能会含有少量的噪声样本，为使超平面不受噪声信号的影响，这里引入松弛变量 ζ。最终，找最优超平面的问题可转化为如下的数学问题：

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{l=1}^{L}\zeta_l \quad (2)$$

$$i_l(w \cdot x_l + b) \ge 1 - \zeta_l,\ \zeta \ge 0,\ l=1,2,\ldots,L \quad (3)$$

式中，C 为惩罚因子，决定对错分样本的惩罚程度，影响分类器对测试样本的鉴别正确率。

上述约束优化问题由Lagrangian函数求解[29-30]。

通过比较线性核函数、多项式核函数、径向基核函数等几种常用核函数的鉴别效果，本文决定采用高斯径向基核函数[30]，其表达式为：

$$K(x_l,x) = \exp\left(-g\|x-x_l\|^2\right),\ g>0 \quad (4)$$

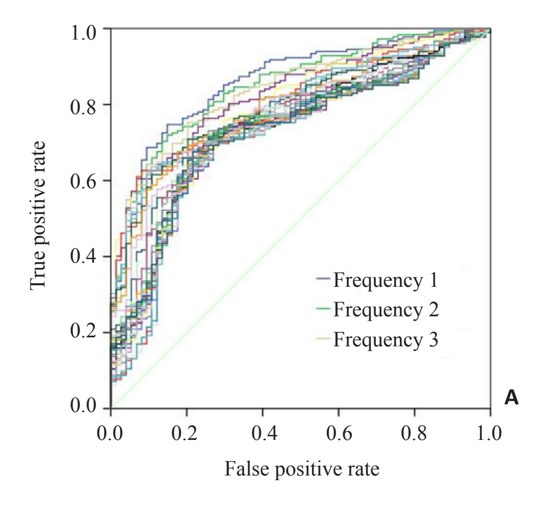
上述惩罚因子 C 和核函数中参数 g 是影响SVM分类器性能的关键参数，需要对其进行寻优。本文采用的参数寻优法为网格搜索法[31]。鉴别正确率采用k折交叉验证[32-33]进行计算。

2 结果

ROC曲线分析和SVM鉴别分别由SPSS 21.0软件和MATLAB R2014a软件来实现，本节将对实验结果进行详细描述。

图2A和B分别展示了所有测量频率点下的介电常数和电导率对应的ROC曲线。图中对角线的AUC为0.5，代表没有鉴别能力。可以看出，所有频率下介电常数和电导率的ROC曲线都完全在对角线上方。图3显示的是图2中所有ROC曲线对应的曲线下面积，所有频率下介电常数和电导率的AUC均在0.7以上，且只有在42.58、53.23、63.87、74.52和85.16 MHz这5个频率下的介电常数的AUC超过0.8。

图4展示了5个Cole-Cole拟合参数对应的ROC曲



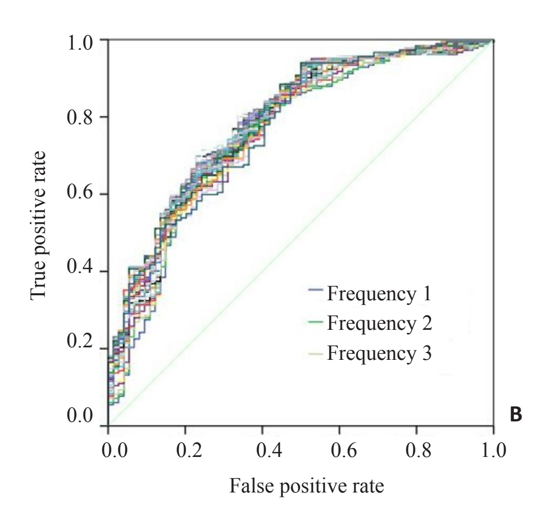


图2 所有测量频率点下的介电常数和电导率对应的ROC曲线

Fig.2 ROC curves of permittivity (*A*) and conductivity (*B*) at all the measured frequency points.

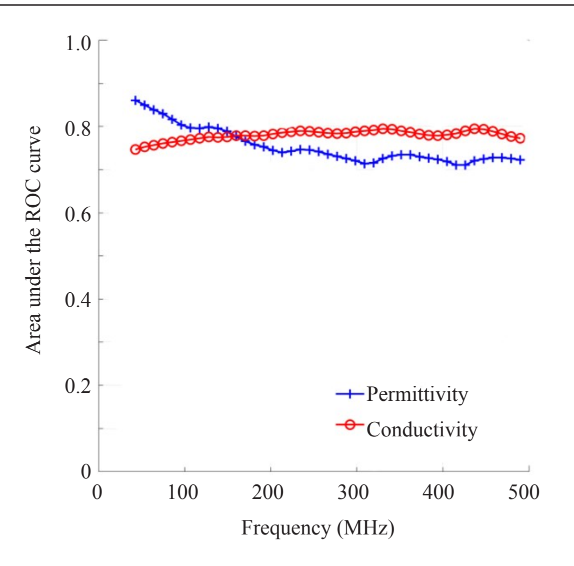


图3 所有测量频率点下的介电常数和电导率对应的ROC曲线下面积

Fig.3 Area under the ROC curves of permittivity and conductivity at all the measured frequency points.

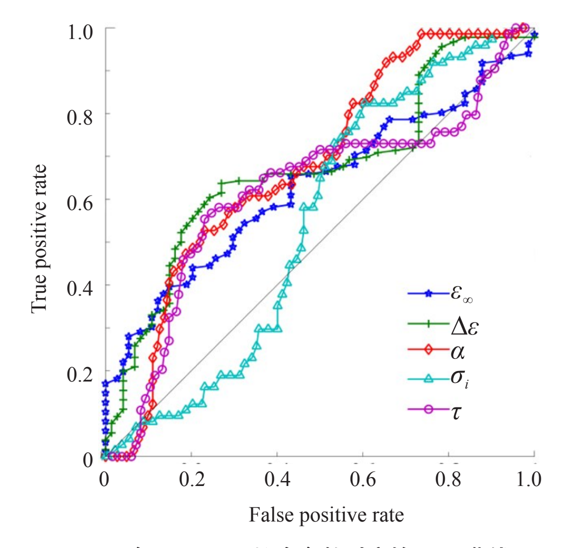


图4 5个Cole-Cole拟合参数对应的ROC曲线

Fig.4 ROC curves for 5 Cole-Cole fitting parameters.

线，可以看出5条曲线均在对角线附近，且参数 ε_∞ 、$\Delta\varepsilon$ 、τ 、α 和 σ_i 各自对应的ROC曲线下面积为0.629、0.671、0.616、0.676和0.544。

根据以上ROC曲线分析的结果，我们选取AUC超过0.8的5个频率下的介电常数(依次记作 ε_{f1} 、ε_{f2} 、ε_{f3} 、ε_{f4} 和 ε_{f5})及其组合作为分类特征值，结合SVM分类器来鉴别正常和恶性的胃组织。以5个频率下介电常数的组合为例：

使用传统的网格搜索法：设定C和g均在[2^{-10}, 2^{10}]范围内，以步距为0.5进行寻优，耗时14.88 s。使用改进的网格搜索法：首先粗略寻优，设定C和g的寻优范围均为[2^{-10}, 2^{10}]，步距均为2，得到局部最优参数组合(C，g)为(256，1)，对应5折交叉验证正确率为82.81%；然后进行精细寻优，在得到的局部最优参数附近，重新定义搜索范围和步距，其中C的取值范围为[2^{5}, 2^{10}]，步距为0.5，g的取值范围为[2^{-3}, 2^{3}]，步距也为0.5，得到最优参数组合(C，g)为(362.0387，8)，对应5折交叉验证正确率为84.38%。整个过程耗时3.40 s。图5(A)和(B)分别展示了粗略寻优和精细寻优的结果图。

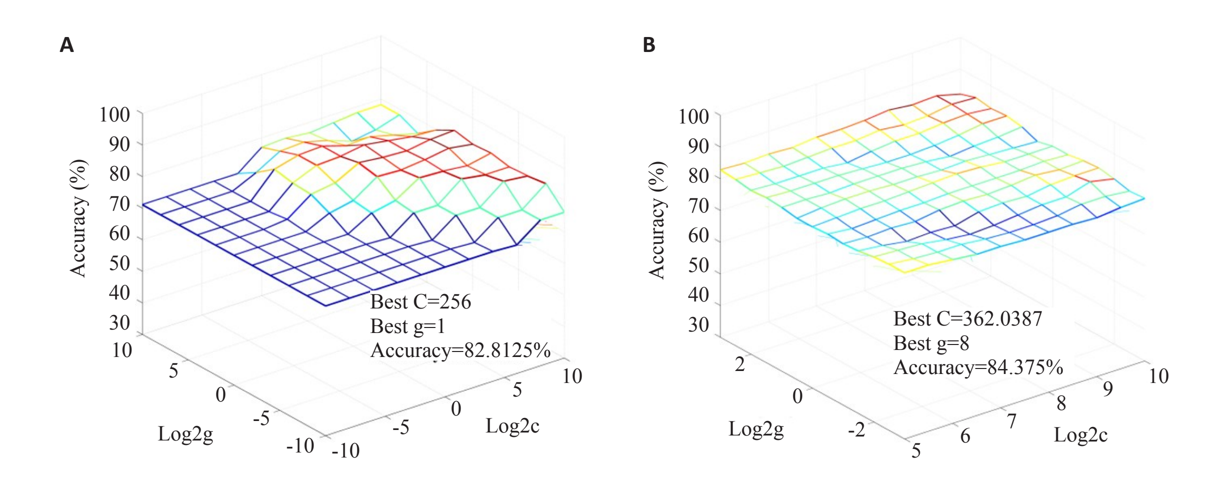


图5 SVM参数选择结果

Fig.5 Results of SVM parameter selection. *A*: Rough search results with grid search method; *B*: Fine search results with grid search method.

表1展示了不同参数组合对应的SVM鉴别正确率和在MATLAB中的运行时间。同一参数组合对应的传统网格搜索法和改进网格搜索法的鉴别正确率一致；改进网格搜索法的运行时间低于传统网格搜索法；5个频率下介电常数组合的鉴别正确率最高，且相对于单个频率下的介电常数并没有增加MATLAB运行时间。

表1 不同参数组合下SVM鉴别结果

Tab.1 SVM discrimination results based on combination of different parameters

Combination of parameters	Traditional grid search		Improved grid search	
	Accuracy (%)	Time (s)	Accuracy (%)	Time (s)
ε_{f1}	81.64	27.96	81.64	3.50
ε_{f2}	79.69	25.26	79.69	3.43
ε_{f3}	78.13	31.15	78.13	3.63
ε_{f4}	77.34	29.11	77.34	4.02
ε_{f5}	75	24.47	75	3.36
$(\varepsilon_{f1}, \varepsilon_{f2}, \varepsilon_{f3}, \varepsilon_{f4}, \varepsilon_{f5})$	84.38	14.88	84.38	3.40

3 讨论

生物组织的介电特性值随着生理或病理状态的改变而改变[34]，目前已有大量文献报道了对正常和恶性生物组织介电特性值的测量，并且证明了二者具有显著差异[8-12, 35-36]，但是并没有具体分析不同频率点下的介电常数、电导率以及Cole-Cole拟合参数的鉴别能力。近几年有少数文献开始基于介电特性用机器学习算法来对正常和恶性生物组织进行鉴别[13-15]，但是目前尚未见文献用该方法对恶性人体胃组织进行自动辅助诊断。本文在测量人体正常和恶性胃组织介电特性的基础上，运用ROC曲线分析各参数的诊断能力，并选出诊断能力较高的参数，同支持向量机结合来对恶性胃组织进行辅助诊断鉴别，并从鉴别正确率和MATLAB运行时间上进行了分析。

本研究测量了42.58 MHz~500 MHz频率范围内43个频率点下正常和恶性胃组织的介电常数和电导率，并用Cole-Cole模型对测量数据进行拟合。通过ROC曲线分析各频率下介电常数、电导率，以及5个Cole-Cole拟合参数的鉴别能力。ROC曲线分析结果显示，所有测量频率点下的介电常数和电导率的ROC曲线下面积都大于0.7，而5个Cole-Cole拟合参数对应的ROC曲线下面积都低于0.7，因此Cole-Cole拟合参数的鉴别能力低于介电常数和电导率。另外，对比不同频率点下的介电常数和电导率，发现只有在42.58、53.23、63.87、74.52和85.16 MHz这5个低端频率下介电常数的ROC曲线下面积超过0.8，说明其具有较高的鉴别能力。根据ROC曲线分析结果，最终选择鉴别能力较高的5个频率下的介电常数及其组合作为分类特征值，结合SVM来鉴别正常和恶性胃组织。SVM鉴别结果显示，5个参数的组合$(\varepsilon_{f1}, \varepsilon_{f2}, \varepsilon_{f3}, \varepsilon_{f4}, \varepsilon_{f5})$作为特征值时取得了最高鉴别正确率84.38%，且改进网格搜索法的MATLAB运行时间为3.40 s，相对于5个参数单独作为特征值并没有增加MATLAB运行时间。因此，在今后的研究中，可以尝试通过测量鉴别能力较高的几个频率点或者较小频率范围内的介电常数，结合改进网格搜索法进行参数寻优的SVM来辅助诊断恶性胃组织，从而避免宽频带数据测量以及Cole-Cole拟合操作。这样可以节省大量时间，增加术中实时辅助诊断的可行性。

本文最终的鉴别正确率和MATLAB运行时间表明，基于介电特性的人体恶性胃组织支持向量机辅助诊断作为一种新的重大疾病辅助诊断方法具有可行性。然而由于病人的性别、年龄、肿瘤分期等因素尚未考虑在内，以及离体测量的影响，本文的鉴别正确率还没有达到最高。接下来的研究将在扩大样本数据量的基础上，考虑上述因素，并找到更好的参数组合，从而进一步提高诊断正确率和时效性。

参考文献：

[1] 辛学刚. 人体组织电特性磁共振断层成像(MR EPT)技术进展[J]. 中国生物医学工程学报, 2015, 34(1): 83-90.

[2] Pethig R. Dielectric properties of body tissues[J]. Clin Phys Physiol Meas, 1987, 8(4A): 5-12.

[3] Schwan H P. Electrical properties of tissues and cell suspensions: mechanisms and models[C]. Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE.IEEE, 1994, 1(1): A70-A71.

[4] Sha L, Ward ER, Stroy B. A review of dielectric properties of normal and malignant breast tissue [C]. Southeast Con, 2002. Proceedings IEEE, 2002: 457-62.

[5] Gabriel C, Gabriel S, Corthout E. The dielectric properties of biological tissues: I. Literature survey[J]. Phys Med Biol, 1996, 41 (11): 2231-49.

[6] Gabriel S, Lau R, Gabriel C. The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz-20 GHz[J]. Phys Med Biol, 1996, 41(11): 2251-69.

[7] Gabriel S, Lau R, Gabriel C. The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues [J]. Phys Med Biol, 1996, 41(11): 2271-93.

[8] Smith SR, Foster KR, Wolf GL. Dielectric properties of VX-2 carcinoma versus normal liver tissue[J]. IEEE Trans Biomed Eng, 1986, 33(5): 522-4.

[9] Surowiec AJ, Stuchly SS, Barr JR, et al. Dielectric properties of

breast carcinoma and the surrounding tissues[J]. IEEE Trans Biomed Eng, 1988, 35(4): 257-63.
[10] Joines WT, Zhang Y, Li C, et al.The measured electrical properties of normal and malignant human tissues from 50 to 900 MHz[J]. Med Phys, 1994, 21(4): 547-50.
[11] Haemmerich D, Staelin ST, Tsai JZ, et al. *In vivo* electrical conductivity of hepatic tumours[J]. Physiol Meas, 2003, 24(2): 251-60.
[12] Stauffer PR, Rossetto F, Prakash M, et al. Phantom and animal tissues for modelling the electrical properties of human liver[J]. Int J Hyperthermia, 2003, 19(1): 89-101.
[13] Truong BCQ, Tuan HD, Fitzgerald AJ, et al. A dielectric model of human breast tissue in Terahertz regime[J]. IEEE Trans Biomed Eng, 2015, 62(2): 699-707.
[14] Truong BCQ, Tuan HD, Wallace VP, et al. The potential of the double debye parameters to discriminate between basal cell carcinoma and normal skin[J]. IEEE Trans Terahertz Sci Technol, 2015, 5(6): 990-8.
[15] Yilmaz T, Kılıç MA, Erdoğan M, et al. Machine learning aided diagnosis of hepatic malignancies through *in vivo* dielectric measurements with microwaves[J]. Phys Med Biol, 2016, 61(13): 5089-102.
[16] Li Z, Deng G, Li Z, et al. A large-scale measurement of dielectric properties of normal and malignant colorectal tissues obtained from cancer surgeries at Larmor frequencies[J]. Med Phys, 2016, 43(11): 5991-7.
[17] Bobowski JS, Johnson T. Permittivity measurements of biological samples by an open-ended coaxial line[J]. Prog Electromagn Res B, 2012, 40: 159-83.
[18] Abdilla L, Sammut C, Mangion LZ. Dielectric properties of muscle and liver from 500 MHz-40 GHz[J]. Electromagn Biol Med, 2013, 32(2): 244-52.
[19] C. Gabriel. Compilation of the dielectric properties of body tissues at RF and microwave frequencies[R]. King's Coll London (United Kingdom) Dept of Physics, 1996.
[20] Peyman A, Gabriel C. Cole-Cole parameters for the dielectric properties of porcine tissues as a function of age at microwave frequencies[J]. Phys Med Biol, 2010, 55(15): N413.
[21] Martellosio A, Pasian M, Bozzi M, et al. Dielectric properties characterization from 0.5 to 50 GHz of breast cancer tissues[J]. IEEE Trans Microw Theory Tech, 2017, 65(3): 998-1011.
[22] Grosse C. A program for the ftting of Debye, Cole-Cole, Cole-Davidson, and Havriliak-Negami dispersions to dielectric data[J]. J Colloid Interface Sci, 2014, 419: 102-6.
[23] Fawcett T. ROC graphs: Notes and practical considerations for researchers[J]. Mach Learn, 2004, 31(1): 1-38.
[24] Hoo ZH, Candlish J, Teare D. What is an ROC curve [J]? Emerg Med J, 2017, 34(6): 357-9.
[25] Cortes C, Vapnik V. Support-vector networks[J]. Mach Learn, 1995, 20(3): 273-97.
[26] Burges CJC. A tutorial on support vector machines for pattern recognition[J]. Data Min Knowl Discov, 1998, 2(2): 121-67.
[27] Chen P, Yuan L, He Y, et al. An improved SVM classifier based on double chains quantum genetic algorithm and its application in analogue circuit diagnosis[J]. Neurocomputing, 2016, 211: 202-11.
[28] Gao X, Hou J. An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process [J]. Neurocomputing, 2016, 174: 906-11.
[29] Wu Y, Tang Z, Xu Y, et al. Support vector regression for measuring electromagnetic parameters of magnetic thin-film materials[J]. IEEE Trans Magn, 2007, 43(12): 4071-5.
[30] Phan AV, Le Nguyen M, Bui LT. Feature weighting and SVM parameters optimization based on genetic algorithms for classification problems[J]. Appl Intell, 2017, 46(2): 455-469.
[31] 李爱明, 刘净瑜, 郝冬梅. 基于改进 CSP 算法的运动想象脑电信号识别方法[J]. 中国生物医学工程学报, 2009, 28(2): 161-5.
[32] Kohavi R.A study of cross-validation and bootstrap for accuracy estimation and model selection[C]. IJCAI, 1995, 14(2): 1137-45.
[33] Galdi C, Wechsler H, Cantoni V, et al. Towards demographic categorization using gaze analysis[J]. Pattern Recognit Lett, 2016, 82: 226-31.
[34] Wang H, He Y, Yang M, et al. Dielectric properties of human liver from 10Hz to 100MHz: normal liver, hepatocellular carcinoma, hepatic fibrosis and liver hemangioma[J]. Biomed Mater Eng, 2014, 24(6): 2725-32.
[35] 王洁然, 王化祥, 徐 晓. 人体肺部组织介电特性实验研究[J]. 中国生物医学工程学报, 2013, 32(2): 178-83.
[36] Hesabgar S M, Sadeghi-Naini A, Czarnota G, et al. Dielectric properties of the normal and malignant breast tissues in xenograft mice at low frequencies (100 Hz-1 MHz)[J]. Measurement, 2017, 105: 56-65.